DOCUMENT RESUME

ED 322 493 CS 010 197

AUTHOR Paul, Peter V.; And Others
TITLE Using Three Test Formats To Assess Good and Poor Readers' Word Knowledge. Technical Report No. 509.
INSTITUTION Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Aug 90
CONTRACT G0087-C1001-90
NOTE 21p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Definitions; Grade 4; Intermediate Grades; Reading Research; *Test Format; Test Wiseness; *Vocabulary; *Vocabulary Skills
IDENTIFIERS Knowledge Acquisition

ABSTRACT

A study was conducted to compare the results of three test formats in relation to assessing the depth of vocabulary knowledge of fourth-grade students. Students (117) were tested on their knowledge of the primary and secondary meanings of 44 multimeaning words using multiple-choice, yes/no, and interview formats. Information on the test-taking strategies used by some students on the multiple-choice tests was also examined and analyzed. It was found that there were no differences in the students' breadth of knowledge of the words, but there was wide variability in their depth of knowledge of the words. The findings suggest that the choice of test format depends on the type of information desired. (Five tables of data are included, and 22 references are attached.) (Author/SR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED322493

CENTER FOR THE STUDY OF READING

Technical Report No. 509

USING THREE TEST FORMATS TO ASSESS GOOD AND POOR READERS' WORD KNOWLEDGE

Peter V. Paul
The Ohio State University

Anne C. Stallman
University of Illinois at Urbana-Champaign

Joseph P. O'Rourke
The Ohio State University

August 1990

University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

CS010197

The work upon which this publication was based was supported in part by the Office of Educational Research and Improvement under Cooperative Agreement No. G0087-C1001-90 with the Reading Research and Education Center. The publication does not necessarily reflect the views of the agency supporting the research.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Anderson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

EDITORIAL ADVISORY BOARD
1989-90

James Armstrong
Linda Asmussen
Gerald Arnold
Yahaya Bello
Diane Bottomley
Catherine Burnham
Candace Clark
Michelle Commeyras
John M. Consalvi
Christopher Currie
Irene-Anna Diakidoy
Barbara Hancin
Michael J. Jacobson
Jihn-Chang Jehng
Robert T. Jimenez
Bonnie M. Kerr
Paul W. Kerr
Juan Moran
Keisuke Ohtsuka
Kathy Meyer Reimer
Hua Shu
Anne Stallman
Marty Waggoner
Janelle Weinzierl
Pamela Winsor
Marsha Wise

MANAGING EDITOR
Fran Lehr

MANUSCRIPT PRODUCTION ASSISTANTS
Delores Plowman
Debra Gough

Abstract

A study was conducted to compare the results of three test formats in relation to assessing the depth of vocabulary knowledge of fourth-grade students. Students were tested on their knowledge of the primary and secondary meanings of 44 multimeaning words using multiple-choice, yes/no, and interview formats. Information on the test-taking strategies used by some students on the multiple-choice tests was also examined and analyzed. It was found that there were no differences in the students' breadth of knowledge of the words, but there was wide variability in their depth of knowledge of the words. The findings suggest that the choice of test format depends on the type of information desired.

USING THREE TEST FORMATS TO ASSESS GOOD AND POOR READERS' WORD KNOWLEDGE

Many researchers agree that vocabulary knowledge is important for reading comprehension (Anderson & Freebody, 1985; Paul & O'Rourke, 1988; Stahl, 1986). In fact, vocabulary knowledge is a potent predictor of a person's ability to comprehend written discourse; however, the reasons underlying this close relationship are not clearly understood. There is no question that good readers have an extensive knowledge of words; that is, they know several meanings, nuances, and other related concepts of many words (Johnson & Pearson, 1984; O'Rourke, 1974). On the other hand, students who know only a few words will almost invariably be poor readers (Anderson & Freebody, 1982, 1985).

Vocabulary, or word, knowledge has two major aspects: breadth and depth (Anderson & Freebody, 1985; O'Rourke, 1974). Breadth of word knowledge refers to the number of words for which at least one meaning or nuance is known. Although this aspect has received the most attention in the research literature, depth of word knowledge, which remains largely unexplored, is equally important. Depth refers to the number of meanings and different usages known for individual words, including multiple meanings, figurative usage of words, and homographs (Anderson & Freebody, 1982, 1985; Johnson & Pearson, 1984). A deep understanding of a word requires knowledge and use of many aspects or concepts related to that word across a variety of contexts (Anderson & Freebody, 1985; Nagy, Herman, Anderson, & Pearson, 1984).

A large proportion of an individual's vocabulary knowledge is gained through encountering words through reading (Nagy, Herman, & Anderson, 1985). To derive and acquire new word meanings, readers must integrate information gathered from several different encounters with words. If readers encounter multiple meanings of a word, they may experience comprehension problems and misinterpret authors (McKeown, 1985). This is a special problem for younger and poorer readers, who may not know secondary or less-common meanings, who have little ability to use context cues to derive the meanings, and who tend to lack flexibility in using meanings appropriately even when they know them (Graves, Slater, & Cooke, 1980; Mason, Kniseley, & Kendall, 1979; Zirkelbach & Blakesley, 1985). For example, Mason et al. (1979) found that children tend to choose the primary (most familiar) meanings of words presented on a sentential level even when the context supported other meanings.

The importance of knowing several meanings of a given word cannot be overemphasized. Johnson, Moe, and Baumann (1983) found that more than two-thirds of the words elementary children encounter have multiple meanings. A study of first-grade basal word lists found that almost 90% of those words have multiple meanings (Searls & Kelsius, 1984). Indeed, multimeaning words constitute a large portion of the lexicon in widely used reading materials such as basal readers.

To make the best use of the limited classroom time available for vocabulary instruction, it is important to measure students' existing vocabulary knowledge as accurately and efficiently as possible. Once levels of vocabulary knowledge have been ascertained, teachers can concentrate on helping students master words they do not know. Teachers who teach vocabulary lessons without first assessing the students' existing knowledge run the risk of using valuable classroom time instructing students on words they already know (Gates, 1961, 1962; Stallman, Commeyras, Kerr, Meyer-Reimer, Jimenez, Hartman, & Pearson, 1989).

The formats typically used to assess vocabulary knowledge are multiple choice, constructed answer, interviews, and yes/no tests. The following discussion reviews the advantages and disadvantages of each of these formats.

Multiple choice. Multiple choice is the most widely used format. The results of vocabulary testing with this format are strongly correlated with other measures of linguistic competence such as intelligence testing and reading comprehension (Anderson & Freebody, 1982, 1985). In addition to the economy of scoring, another advantage is that this format can be designed to be sensitive to *partial* knowledge of items (Kolstad, Briggs, Bryant, & Kolstad, 1983; Nagy et al., 1985). For example, Nagy et al. used distractors of varying levels of difficulty to determine how closely the student's knowledge of the meaning of a word came to complete adult-like understanding of the word's meaning. With this procedure, students can be given credit for partial, as well as complete, knowledge of a word's meaning. This is important in vocabulary testing because children often acquire knowledge of words in incremental stages (Anderson & Freebody, 1985; Dale & O'Rourke, 1986). That is, when they learn new words, children may initially understand some aspects of the word and then continue to add to that knowledge through further experiences.

There are two major drawbacks to the multiple-choice format. One is the difficulty of choosing distractors that suppress the possibility of guessing correctly without misdirecting the subject (Anderson & Freebody, 1982, 1985; Kolstad, Briggs, Bryant, & Kolstad, 1983; Nagy et al., 1985). The other is that multiple-choice tests are unusually sensitive to the influence of test-taking strategies. For example, failure of the student to consider all the options may negatively affect the results, especially with poor readers (Anderson & Freebody, 1982, 1985).

Interviews. Student interview techniques effectively solve some of the problems of the multiple-choice format. Interviewing presents no problem of choosing distractors, and individual differences can be taken into account. The interview format is also sensitive to partial knowledge on the part of subjects because it allows the examiner to modify procedures to probe for more complete information. This format, however, is time consuming to administer, and the data are difficult to score reliably (Anderson & Freebody, 1982, 1985; Nagy et al., 1985).

Yes/no. The yes/no format, a procedure that uses a mixture of words and nonwords and a formula to correct for guessing, may be an attractive alternative to both the multiple-choice and interview formats (Anderson & Freebody, 1982, 1985). This procedure makes efficient use of testing time and is sensitive to partial knowledge. It may be a more valid predictor of vocabulary knowledge than the multiple-choice format (using interview scores as the criterion measure for evaluating the concurrent validity of each). The yes/no format, however, cannot be used to determine which meanings of multimeaning words are known, nor can it be used to assess the effectiveness of direct vocabulary instruction because no estimate of individual word knowledge is possible with this format.

The purpose of this study was to compare the results of three test formats in relation to the assessment of depth of vocabulary knowledge of fourth-grade students. Students were tested on their knowledge of the primary and secondary meanings of 44 multimeaning words using multiple-choice, yes/no, and interview formats. Information on the test-taking strategies that some students used on the multiple-choice tests was also obtained and analyzed.

It was hypothesized that good readers, regardless of format, would not only score higher but would also evidence greater depth of word knowledge. It was also hypothesized that all students would perform better on the primary, or most common, meanings of the words. Finally, it was hypothesized that the results would conform to the "hierarchy of meanings" phenomenon reported by Dale and O'Rourke (1976, 1981). In other words, certain meanings of multimeaning words are more likely to be known than others. For example, students who know a secondary meaning of a word also know the primary meaning, and those students who know only one meaning know only the primary meaning of the word.

Method

Subjects

One hundred-seventeen fourth-grade students from four self-contained classrooms in a midwestern public school served as subjects. None of the students had any apparent educational disabilities. Because no standardized test scores were available, the students were divided into high- and low-ability groups based on their placement in the *Ginn Reading Program*, which was used for reading instruction in the school system. Reading grade levels for the high-ability group ranged from 4.6 to 5.7 (mean = 4.9) and from 3.3 to 4.4 (mean = 3.8) for the low-ability group. The groups were balanced for sex, race, and number of years in school.

Of these students, 40 were randomly selected to be interviewed, 20 from the high-ability group and 20 from the low-ability group. These groups were also balanced for sex, race, and number of years in school. For a second set of interviews, five additional students were randomly selected and added to each group.

Instrumentation and Procedures

The selection of the 44 words was based on the following criteria:

1. The primary meaning of each word was known by 85% or more fourth graders according to norms in *The Living Word Vocabulary* (Dale & O'Rourke, 1976, 1981). This was done in order to select words for which most of the students would know at least one meaning.

2. There was a difference of at least seven percentage points between the percent of fourth graders knowing the primary and secondary meanings of each word (Dale & O'Rourke, 1976, 1981).

3. No homographs were used.

4. All the words fell within the 15,000 most frequent words (Carroll, Davies, & Richman, 1971). This was done in order to eliminate words that are likely to be unknown by the students simply because they occur infrequently.

The 44 words were assessed on a yes/no test. Pseudo words were also used and were randomly mixed in with the real words. For example, *blink* was changed to *blonk*. Pseudo words and real words were matched in terms of length and difficulty. The students were asked to read each word silently, circle it if the meaning was known, and cross it out if the meaning was not known. The yes/no test was administered first so that students would not mark the words as known simply because they had seen them on one of the other tests.

Two multiple-choice tests (Dale & O'Rourke, 1976, 1981) were used to test students' knowledge of word meanings. One test assessed knowledge of the primary meanings of the words, and the other assessed knowledge of the secondary meanings of the same words. Each item consisted of the target word followed by three options. The correct answers consisted of synonyms or short phrases that captured the particular meaning of the word being tested. The distractors were either semantically or phonologically similar to the target word.

The multiple-choice tests were administered to the students in their normal classroom settings. Four different forms (i.e., different random orders of items) of each test were used to reduce the possibility

of students' copying from one another and to control for order effects. The multiple-choice test on the secondary meanings was administered before the test on the primary meanings because, presumably, the secondary meanings are more difficult, and it was possible that exposure to the primary, or easier, meanings might create interference in accessing the secondary meanings later. The students were given game activities between the tests to reduce the likelihood of carryover effects. The reliability of the multiple-choice tests according to the Kuder-Richardson (K-R) 20 procedure was quite high, .92 for the primary meanings test and .80 for the secondary meanings test.

The interviews were conducted individually in separate rooms. During the first set of interviews, students were shown each word, asked to pronounce it, and to give its meaning. If the students could not give a definition of the word, they were asked to use it in a sentence or to tell something about the word. If students provided any of the above information, they were then asked to give another meaning, sentence, or example. Pronunciation of words was corrected or provided for those students who were unable to pronounce the words. Students were also asked to define the pseudo words they had indicated they knew on the yes/no tests. Because it is possible that taking the multiple-choice tests can influence the performance in the interviews, the interviews were conducted several weeks after the students had taken the multiple-choice tests to minimize any carryover effects.

A second set of interviews was conducted to examine the test-taking strategies used on the multiple-choice tests. Students were shown a blank copy of the multiple-choice test they had originally taken for the secondary meanings of the words. The 20 items used in these interviews were based on two criteria. First, students were asked to respond to the 11 words for which *no* student had given the secondary meaning tested during the first set of interviews. Second, the other 9 words were selected for the students individually. The students were encouraged to think out loud as they worked. These interviews were conducted to examine the types of test-taking strategies used on multiple-choice tests. The second set of interviews was conducted about two months after the paper and pencil tests and about three weeks after the first set of interviews.

Results and Discussion

Using correlational techniques, this study compared three different test formats--multiple-choice, yes/no, and interview--to determine the similarity in informational yield. If the results indicated that the three measures are highly similar, then group multiple-choice or yes/no measures could be efficient substitutes for the more timeconsuming interview format. Assessment of good and poor readers' knowledge of multimeaning words was also of interest. To determine differences in word knowledge of primary versus secondary meanings, scores on the primary-meaning test were compared to scores on the secondary-meaning test using ANOVA procedures. A chi-square analysis was performed to determine whether a hierarchy of meanings existed. Finally, the researchers explored the test-taking strategies of the students on the multiple-choice tests. The strategies were examined in terms of similarities and differences between good and poor readers. The strategy analysis also helped determine the accuracy of the students' scores on the multiple-choice tests (i.e., whether students really knew the meaning of the word or whether they guessed).

Differences Attributable to Ability or Test Format

As expected, good readers performed better than poor readers on all three formats (see Table 1). Analysis of variance was performed on the multiple-choice format to examine differences between the groups and differences between the two multiple-choice tests. Group differences and test differences were both significant beyond the .0001 level, but the Group X Test interaction was not significant (see Table 2). This indicates that good readers know more meanings of words than poor readers and that all students are more likely to know primary better than secondary meanings.

[Insert Tables 1 and 2 about here.]

During the first set of interviews, students were required to state what they knew about all words tested on the multiple-choice and yes/no formats. The interview scores were correlated with the scores on each multiple-choice test separately and on the yes/no test to determine which of the two paper-and-pencil formats best predicted interview performance (see Table 3). The scores on the yes/no tests were correlated with the words from the interviews for which the student gave *any* meaning that was correct, whether the meaning the student gave was one that was being tested on the multiple-choice test or not. All correlations were statistically significant ($p < .05$). These results indicate that the multiple-choice tests are representative of the knowledge that students have of specific meanings of words and that the yes/no test is a good indicator of whether or not the students have *any* knowledge of the words.

[Insert Table 3 about here.]

The fact that, overall, the students performed better on the primary meanings of the words than they did on the secondary meanings seems to support a hierarchy of meanings for words. A chi-square analysis of the number of students who knew the two tested meanings of the words indicated that a significant number of students knew one of the meanings more frequently than the other for 64% of the words (see Table 4). In addition, when students knew only one meaning of a word they were more than three times as likely to know the primary meaning. In respect to other words, there were a few cases, such as the word *punch*, in which all the students knew both tested meanings. For others, such as *clip*, virtually none of the students knew either tested meaning.

[Insert Table 4 about here.]

In general, the results conformed to those of Dale and O'Rourke (1976, 1981). Discrepancies between results can be attributed to factors such as the fluid nature of language and differences in the students' backgrounds. Language usage changes over time and word meanings that are commonly used by one generation may be slightly out of vogue for the next. There were several words for which this seems to be the case. For example, the primary meaning for the word *tube* in the *Living Word Vocabulary* is "used in TV set." None of the students in the present study gave that meaning for *tube*. However, they did know the secondary meaning, "long, hollow container." Words that fall into this category are responsible for most situations in which students knew the secondary, but not the primary meanings. In these cases it seems logical to reassign the listed secondary meanings as primary and the primary as secondary. This procedure indicated that, when the students knew only one meaning of a word, they were more than ten times as likely to know the primary meaning of the word--as opposed to three times as likely using the original categorization scheme. It should also be noted that, if we had used the second categorization scheme for all of our analyses, those results would have been similarly enhanced.

Advantages and Disadvantages of the Three Test Formats

The interviews involved a task of recall. What was recalled was influenced by a variety of factors such as the topics currently being discussed in school or the time of the year. For example, the testing was conducted in mid-February and all students mentioned valentines in connection with the word *heart*, but failed to mention a parade vehicle for *float*. This trend might have been reversed if students had been tested in early July. Thus, one limitation of the interview format is the possibility that it may not tap all knowledge students really possess. This was especially evident for the high-ability group. Their scores on the multiple-choice tests (in which they had to *recognize* or *identify* meanings) indicated that they knew more word meanings than they were able to demonstrate in an interview format. Furthermore, these multiple-choice data are rendered more credible by the strategy interviews, which indicated that the students were not guessing on the multiple-choice tests (see Table 5). The main

advantage of the interview format is that it provides greater opportunities for the examiner to gain insights into how students approach and figure out answers to questions.

Because the interview format is time consuming to administer (about 30 minutes per student), paper-and-pencil measures seem to be more efficient in obtaining certain kinds of information (about 15 minutes per group of students). Paper-and-pencil tasks, such as multiple-choice and yes/no tests, yield different kinds of information about students' vocabulary knowledge.

The yes/no test is useful for testing breadth of vocabulary knowledge. A large number of words can be tested in a relatively short period of time, and the results indicate the number of words about which students have some knowledge. However, with this type of test, there is no way of ascertaining *what* students know about the words or *which* words they know. Knowledge of individual words can range from simply recognizing that a particular word is a real English word to knowing all the meanings and nuances that are associated with the word.

Another limitation of the yes/no test is the formula that is used to score results:

$$\frac{P(H) - P(FA)}{1 - P(FA)}$$

The formula indicates that the percentage of real words to which a person says "yes," P(H) (literally the proportion of hits), minus the proportion of nonwords to which a person says "yes," P(FA) (literally the proportion of false alarms), divided by one minus the proportion of nonwords to which a person says "yes" (Anderson & Freebody, 1982). This formula works in all cases except the one in which a person says "yes" to all real words. In this case, a score of 100% is obtained, regardless of the percentage of nonwords to which a person says "yes." For example, a person who says "yes" to all real words and none of the nonwords receives the same score as the one who says "yes" to all real words and half of the nonwords. While both individuals receive a score of 100%, the first one evidences a higher degree of vocabulary knowledge. Ultimately, some change in the formula is required.

To determine *what* a person knows about a word or whether a person knows any *particular* word, it is necessary to use a format such as a multiple-choice test. The multiple-choice format is an efficient means of assessing *depth* of vocabulary knowledge. It can be used to assess knowledge of specific word meanings and is far less time-consuming to administer and score than constructed answer or interview formats.

Test-Taking Strategies

The purpose of the second set of interviews was to obtain information on strategies that students used on a multiple-choice test and the success rate of the various strategies (see Table 5). The observed strategies were categorized as readability of the options, position of the options, elimination, association, guessing, and knowing the answer. These strategies were inferred from reasons that students offered for their choice of responses during the interviews. For example, when students said that a response was chosen because it was the only one they could read, that strategy was labeled "readability of the options." The strategy of choosing a response because it was first, last, or in the middle was labeled "position of the options." The strategy that involved the selection of an option after ruling out all others for various reasons was labeled "elimination." The "association" strategy involved the selection of an option because it could be related in some way to something students knew about the word. Guessing strategies involved one of two situations: (a) students didn't know why they selected a response, or (b) students said they guessed. The final category, "knowing the answer," was used when students stated that they chose the response because they knew it was right.

[Insert Table 5 about here.]

The most frequently used strategy (54% of the time) of students in both the high-ability and low-ability groups was association. Both groups used it successfully a little over half of the time. Sometimes the students did not know the meaning for a particular word. In order to come up with a response, they tried to associate something else they knew about that word using the options from which they had to choose. For example, they chose "a group" as the correct response for *join* because "you can join a group." For the remaining strategies, the two groups differed in frequency of use and success rate.

The high group used strategies in the following frequency order: association (54% of the time), elimination (19%), knowing the answer (16%), guessing (8%), and position of the options (3%). Knowing the answer was the most successful strategy and resulted in correct responses 97% of the time. Elimination resulted in correct responses 64% of the time. The other strategies, guessing and position of the options, were basically unsuccessful strategies and were rarely used.

The students in the low group used strategies in the following frequency order: association (54% of the time), guessing (21%), elimination (13%), knowing the answer (8%), readability of the options (3%), and the position of the options (.6%). Again, the most successful strategy (93% correct) was knowing the answer; but, this strategy was rarely used. The success rate of the elimination strategy of the low group was similar to that of the high group. However, the other strategies used by the low group differed radically from the high group both in frequency of use and success rate. For example, the low-group students resorted to guessing almost three times more frequently than students in the high group and were much less successful. Another unsuccessful strategy was reliance on the readability of the options. While this strategy was used relatively infrequently by students in the low group, it was not employed at all by students in the high group. Overall, other than guessing, the success rate of the strategies was similar for the two groups, the difference lies in the frequency with which the strategies were used.

Conclusion

Vocabulary knowledge can be considered in terms of breadth and depth, and both types of knowledge are essential for effective reading comprehension. To increase the vocabulary knowledge of students, the first step is to measure the knowledge they already possess. Several different formats are typically used to measure word knowledge. The choice of format depends on the type of information desired and time constraints imposed by the setting.

The interview format may be the most effective way to find out exactly what students know about specific words. However, this format is time consuming to administer and difficult to score reliably. Thus, it is not practical for testing more than a few students or a large number of words.

An efficient way to test students' breadth of vocabulary knowledge is with a yes/no format. This format lends itself to the testing of groups of students on large numbers of words in a relatively short period of time, but it does not provide any information on what students know about words or even which words they know something about. The information from the yes/no format can be used to rank students according to their breadth of vocabulary knowledge. Nevertheless, in this study, there were no differences in the high and low groups' breadth of vocabulary knowledge, but there was wide variability in the groups' depth of knowledge of the words.

The multiple-choice format is an efficient way to assess depth of vocabulary knowledge because it can be administered to groups of students and is fairly easy to score. Because it can be used to determine which meanings of words are known by students, it has clear implications for classroom use. Once

students' word knowledge has been determined, the assessment data can be used to help teachers concentrate instructional time on the meanings of words that students do not know.

References

Anderson, R., & Freebody, P. (1982). *Reading comprehension and the assessment and acquisition of word knowledge* (Tech. Rep. No. 249). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Anderson, R., & Freebody, P. (1985). Vocabulary knowledge. In H. Singer & R. Ruddell (Eds.), *Theoretical models and processes of reading* (3rd ed., pp. 343-371). Newark, DE: International Reading Association.

Carroll, J., Davies, P., & Richman, B. (1971). *The American heritage word frequency book*. New York: American Heritage.

Dale, E., & O'Rourke, J. (1976). *The living word vocabulary: A national vocabulary inventory*. Elgin, IL: Field Enterprises.

Dale, E., & O'Rourke, J. (1981). *The living word vocabulary: A national vocabulary inventory* (2nd ed.). Chicago, IL: World Book.

Dale, E., & O'Rourke, J. (1986). *Vocabulary building: A process approach*. Columbus, OH: Zaner-Bloser.

Gates, A. (1961). Vocabulary control in basal reading material. *Reading Teacher, 15,* 81-84.

Gates, A. (1962). The word recognition ability and the reading vocabulary of second- and third-grade children. *Reading Teacher, 15,* 443-448.

Graves, M., Slater, N., & Cooke, C. (1980, December). *Skill with polysemous words: A measurement of the depth of children's word knowledge*. Paper presented at the annual meeting of the National Reading Conference, San Diego.

Johnson, D., Moe, A., & Baumann, J. (1983). *The Ginn word book for teachers: A basic lexicon*. Lexington, MA: Ginn.

Johnson, D., & Pearson, P. D. (1984). *Teaching reading vocabulary*. New York: Holt, Rinehart & Winston.

Kolstad, R., Briggs, L., Bryant, B., & Kolstad, R. (1983). Complex multiple-choice items fail to measure achievement. *Journal of Research and Development in Education, 17,* 7-11.

Mason, J., Kniseley, E., & Kendall, J. (1979). Effects of polysemous words on sentence comprehension. *Reading Research Quarterly, 15,* 49-65.

McKeown, M. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly, 20,* 482-496.

Nagy, W., Herman, P., & Anderson, R. (1985). Learning words from context. *Reading Research Quarterly, 20,* 233-253.

Nagy, W., Herman, P., Anderson, R., & Pearson, P. D. (1984). *Learning words from context* (Tech. Rep. No. 319). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

O'Rourke, J. (1974). *Toward a science of vocabulary development.* The Hague: Mouton.

Paul, P., & O'Rourke, J. (1988). Multimeaning words and reading comprehension: Implications for special education students. *Remedial and Special Education* (RASE), *9*(3), 42-52.

Searls, E., & Klesius, J. (1984). 99 multiple meaning words for primary students and ways to teach them. *Reading Psychology, 5,* 55-64.

Stahl, S. (1986). Three principles of effective vocabulary instruction. *Journal of Reading, 29,* 662-668.

Stallman, A., Commeyras, M., Kerr, B., Meyer-Reimer, K., Jimenez, R., Hartman, D., & Pearson, P. D. (1989). *Are "new" words really new?* (Tech. Rep. No. 471). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Zirkelbach, T., & Blakesley, K. (1985). The language deficient child in the classroom. *Academic Therapy, 20,* 605-612.

Table 1

Mean Scores (% correct) and Standard Deviations by Group for Each Test Format

Total Group (n = 117)

	Multiple-choice Tests		Yes/No Tests
	Primary	Secondary	
Low Group	81.18 (17.84)	66.98 (13.98)	85.12 (16.68)
High Group	96.07 (9.39)	80.70 (9.48)	96.97 (4.41)

Interviewed Students (n = 40)

	Multiple-choice Tests		Yes/No Tests
	Primary	Secondary	
Low Group	77.85 (18.11)	67.00 (13.76)	79.00 (21.06)
High Group	93.91 (6.13)	78.41 (6.45)	96.05 (4.02)

Interviews

	Primary	Secondary	Any
Low Group	69.50 (10.75)	32.95 (9.99)	88.80 (12.03)
High Group	76.09 (5.00)	41.79 (7.76)	97.45 (2.24)

Table 2

ANOVA of Multiple-choice Test Scores on Primary and Secondary Word Meanings by Ability Groups

Source	df	Sum of Squares	F
Group	1	2318.01	42.59*
Subject Within Group	115	6258.52	4.98*
Test	1	2479.56	227.05*
Group X Test	1	3.72	.34
Error	115	1255.90	
Total	233	12317.54	

*$p < .0001$

Table 3

Correlations Between Paper/Pencil Test Scores and Interview Scores

Total Group (n = 40)

	Reading Level	Interview I (Primary)	Interview I (Secondary)	Interview I Any
Reading Level	------	.51*	.51*	.55*
Multiple-choice Test (Primary)	.60*	.74*	.66*	.82*
Multiple-choice Test (Secondary)	.57*	.73*	.69*	.77*
Yes/No Test	.55*	.69*	.71*	.81*

*$p < .05$

Table 4

Students' Knowledge of Primary (1st) and Secondary (2nd) Word Meanings

Word	1st-yes 2nd-yes	1st-yes 2nd-no	1st-no 2nd-no	1st-no 2nd-yes
account*	1	27	12	0
allowance*	0	39	1	0
beach*	2	0	1	37
bend	6	7	1	26
blink*	6	0	0	34
blue*	16	24	0	0
bottle*	1	39	0	0
bowl	31	9	0	0
branch*	0	40	0	0
build*	4	36	0	0
bump	12	6	3	10
bunk*	4	4	4	28
chest	23	17	0	0
clip	0	0	32	8
complete*	4	35	0	1
deck	5	10	15	10
dirty*	11	29	0	0
draw	25	15	0	0
engineer*	0	16	24	0

Table 4 (Continued)

Word	1st-yes 2nd-yes	1st-yes 2nd-no	1st-no 2nd-no	1st-no 2nd-yes
file*	8	1	5	26
float*	12	28	0	0
head*	1	39	0	0
heart*	2	38	0	0
join	15	14	2	9
key*	0	40	0	0
loaf*	10	29	1	0
match*	3	31	4	0
medicine*	0	40	0	0
note*	1	39	0	0
part*	9	2	4	25
post	1	16	18	5
punch	39	1	0	0
salt*	0	40	0	0
side	4	18	4	14
skin*	9	31	0	0
slip*	1	35	2	2
sneak	13	14	5	12
stick	22	15	2	1
straight*	2	35	0	3

Table 4 (Continued)

Word	1st-yes 2nd-yes	1st-yes 2nd-no	1st-no 2nd-no	1st-no 2nd-yes
stuck	3	24	1	12
sub	28	10	1	1
suit*	3	35	2	0
tape	32	5	1	2
tube*	2	0	5	31
Total	371	935	146	308

*Chi-square analysis of student's knowledge of primary versus secondary word meanings, $p < .05$.

Table 5

Test-Taking Strategies

Strategy	High Group *n*	High Group % Total	High Group % Success	Low Group *n*	Low Group % Total	Low Group % Success
Readability of options	0	0	--	14	3	14
Position of options	15	3	40	3	6	67
Elimination	96	19	64	67	13	61
Association	272	54	61	269	54	57
Guessing	39	8	51	106	21	35
Knowing the answer	78	16	97	41	8	93
Total	500	100	66	500	100	54.8